doi: 10.2514/1.I010772

Identification of Instantaneous Anomalies in General Aviation Operations using Energy Metrics

Tejas G. Puranik * and Dimitri N. Mavris †
Georgia Institute of Technology, Atlanta, GA 30332-0150

Quantification and improvement of safety is one of the most important objectives among the General Aviation community. In recent years, machine learning techniques have emerged as an important enabler in the data-driven safety enhancement of aviation operations with a number of techniques being applied to flight data to identify and isolate anomalous (and potentially unsafe) operations. Energy-based metrics provide measurable indications of the energy state of the aircraft and can be viewed as an objective currency to evaluate various safety-critical conditions across a heterogeneous fleet of aircraft and operations. In this paper, a novel method of identifying instantaneous anomalies for retrospective safety analysis in General Aviation using energy-based metrics is proposed. Each flight data record is processed by a sliding window across the multi-variate time series of evaluated metrics. A Gaussian Mixture Model using energy metrics and their variability within each window is fit in order to predict the probability of any instant during the flight being nominal. Instances during flights that deviate from the nominal are isolated to identify potential increased levels of risk. The identified anomalies are compared with traditional methods of safety assessment such as exceedance detection to highlight the benefits of the developed method. The methodology is demonstrated using flight data records from two representative aircraft for critical phases of flight.

Nomenclature

h	=	Altitude Above Touchdown or Takeoff (m)
V	=	Velocity (m/s)
g	=	Acceleration due to gravity (m/s^2)
$\dot{V}$	=	Acceleration (m/s^2)
T	=	Thrust (Newton)
D	=	Drag (Newton)
W	=	Weight of the aircraft (Newton)
γ	=	Flight path angle (degree)
h_{ref}	=	Altitude for Reference Profile for Phase of Flight (m)
V_{ref}	=	Velocity for Reference Profile for Phase of Flight (m/s)
γ_{ref}	=	Flight Path Angle for Reference Profile for Phase of Flight (m)
T_{max}	=	Maximum Thrust Available (Newton)
T_{idle}	=	Idle Thrust (Newton)

I. Introduction

Aviation is statistically one of the safest modes of transportation within the US [1] and with a steadily improving safety record over the past few decades [2]. While the overall safety record in aviation has improved, despite the best efforts of regulators and safety practitioners, General Aviation (GA) safety continues to lag behind commercial aviation. A majority of accidents that occur in aviation are in the GA domain according to the National Transportation Safety

*Research Engineer II, Aerospace Systems Design Laboratory, Daniel Guggenheim School of Aerospace Engineering, AIAA Member
†S.P. Langley NIA Distinguished Regents Professor, Director of Aerospace Systems Design Laboratory, Daniel Guggenheim School of Aerospace Engineering, AIAA Fellow

Board (NTSB) [1] statistics. Therefore, improving GA safety has been among the top priorities for various regulatory bodies in recent years [3]. According to the Federal Aviation Administration (FAA), the demand for air travel and traffic is predicted to grow steadily over the next two decades at a rate of approximately 1.8% annually [4]. Commercial operations are expected to double and GA operations are also expected to gain a much-needed revitalization in the coming years with the advent of urban air mobility and other novel concepts. With such a large increase in expected operations, there is an ever-increasing demand for improving safety of all aviation operations, especially GA.

In the past, accidents have been the primary triggers for identifying problems and developing mitigation strategies [5]. However, the industry has moved towards a more proactive approach to safety enhancement in which potential unsafe events are identified beforehand and mitigation strategies are implemented in order to prevent accidents. The FAA have implemented various safety improvement programs such as the Flight Operational Quality Assurance (FOQA) [6], Aviation Safety Action Program (ASAP) and others. Aviation Safety Information Analysis and Sharing (ASIAS) [7] is a system which aims to connect a number of such safety databases in order to facilitate integrated queries across multiple databases. While many of these and other programs initially started for commercial operations where more data is available, recently, they have trickled into GA as well. Specifically within GA, the General Aviation Joint Steering Committee (GAJSC) is a government and industry group that analyzes GA safety data to develop intervention strategies to prevent or mitigate problems associated with accident causes, called Safety Enhancements (SE) [8].

The most common and widespread programs in existence for quantitative safety assessment are Flight Operational Quality Assurance (FOQA) [6] or Flight Data Monitoring (FDM) [9]. It is defined as: "*The systematic proactive use of digital flight data from routine operations to improve aviation safety within a non-punitive and just safety culture*". Retrospective analysis of flight data in FOQA programs is one of the most important enablers for quantitative safety assessment. FOQA programs aim to improve operational safety with a continuous cycle involving data collection from on-board recorders, retrospective analysis of flight data records, identification of operational safety exceedances, design and implementation of corrective measures, and monitoring to assess their effectiveness. Implementation of FOQA is widespread in commercial aviation, and although it is sparse among GA operators, recent and current efforts seek its introduction and broad adoption in GA [10].

Exceedance detection is the most common method of analysis using flight data. An exceedance is the deviation of a single parameter beyond an established threshold. An event is defined by one or more parameter exceedances that take place concurrently over a specified period of time. The current practice performs well on known safety issues but is blind to safety-critical conditions that may be captured by flight data records but not included in the set of predefined safety events. Additionally, the events defined in exceedance detection are 'static' in that they do not take into account the information that might be available in a wealth of recorded flight data. With the availability of flight data and advanced computing power, data mining and machine learning techniques for safety analysis, incident examination, and fault detection have garnered increased interest in the aviation community.

While the accident rates for GA flight operations are higher than those in commercial aviation, it is noted that GA contains a heterogeneous fleet and operations. This includes single and multi-engine piston, turboprop, turbojet powered aircraft as well as helicopters and experimental aircraft. However, accident and incident rates and number of active aircraft are not uniform across all the aircraft classes within GA. Historically single-engine piston aircraft make up a significant proportion of the entire fleet and the number of accidents and fatalities [11]. Therefore, it is of particular interest to examine operations of this class of aircraft within GA.

These aircraft typically belong to the *normal* category under the FAA airplane certification categories (14 CFR §23.2005 [12]) which applies to airplanes with a passenger-seating configuration of 19 or less and a maximum certificated takeoff weight of 19,000 pounds or less. In this category, airplanes within level 1 (maximum seating configuration of 0 to 1 passengers) and level 2 (maximum seating configuration of 2 to 6 passengers) of airplane certification levels are of particular interest.*

More exhaustively, the following important points distinguish the category of aircraft and operations that are the focus of this paper:

1) Smaller sized aircraft
2) Less weather-tolerant aircraft
3) Limited or no flight data recording capability
4) Highly variable and heterogeneous mission profiles
5) Variability in pilot certificate and experience level (number of hours flown)
6) Greater variety of airports of operations

*It is worth noting that the recent rewrite to 14 CFR §23 has changed the aircraft categories and broke it out into six airplane certification levels and two airplane performance levels. The reader is directed to reference [12] for more details.

7) Operations mostly under Visual Flight Rules (VFR)

All of these factors present important challenges for improving safety of GA operations. Therefore, existing mature methods from commercial operations need to be tailored to GA operations or entirely new methods for analysis of GA data need to be developed. The rest of the paper is organized as follows – Section II contains a review of relevant past work related to the identification of anomalies from flight data and the research objective for the current work. Section III presents the details of the methodology developed for identifying instantaneous anomalies. Section IV contains the results obtained from the application of this methodology on a set of real flight data from GA operations from two representative single-engine aircraft and the comparison of these results to traditional exceedance detection techniques. Section V provides the conclusions and recommendations from this body of work and outlines avenues for future work.

II. Background and Research Objective

Previous applications of data mining in the aviation safety domain have primarily treated it as an anomaly detection problem with data objects as multivariate time series [13–15]. In the data mining community, anomaly detection is defined as the *"task of obtaining patterns in data that do not conform to a well defined notion of normal behavior"* [16]. The exact notion of anomaly is different for different application domains. In the aviation safety domain, two main types of in-flight anomalies are of interest to the safety analyst – *instantaneous* and *flight-level*. In addition to these, various pre-flight checks are also typically conducted to find system faults or anomalies prior to takeoff. Instantaneous anomalies refer to those anomalies where only an instant or small part of the flight record is abnormal. Flight level anomalies refer to those flights which exhibit abnormal data patterns over an entire flight or phase of flight. While both types of anomalies are important and offer different insight into the problem, the main focus of this paper is identification of instantaneous anomalies in GA flight data. Within each type of anomaly, three types of techniques are observed – supervised, semi-supervised, and unsupervised techniques. Supervised techniques rely on a labeled training set consisting of typical system behaviors as well as anomalous behaviors to train algorithms which can then be used on new flight data. Semi-supervised techniques only require a set of training data that contains mostly nominal behaviors. Finally, unsupervised techniques operate under the assumption that the given data set may contain anomalous as well as normal operations in any proportion.

A. Previous Work

The main objective of anomaly detection techniques is to detect abnormal flights within routine flight data without any prior knowledge of safety events (unsupervised or semi-supervised). They involve the use of various machine learning algorithms (clustering, classification etc.) on flight data or accident data to identify emerging risks. Typically, the data obtained from flights is pre-processed to generate features that can be used in the algorithms. These features are generated from the recorded parameters using the temporal nature of the data. Anomalous flights obtained (sometimes as ordered lists based on their 'anomalousness') are then further analyzed by subject matter experts. Gavrilovski et al. [17] have provided a comprehensive survey of data-mining and anomaly detection techniques applied to flight data. Some of the seminal work in the application of anomaly detection techniques to flight data is summarized here.

There have been several applications of data mining techniques to aviation data focusing on flight level anomalies. Das et. al. [13] have developed Multiple Kernel Anomaly Detection (MKAD) which applies a one-class support vector machine for flight level anomaly detection. MKAD identifies flight level anomalies well in heterogeneous data but is not built to identify instantaneous anomalies. Matthews et al. [14] have discussed and summarized the aviation knowledge discovery pipeline using various algorithms and attempted to combine the strengths of different approaches. Li et al. [15] have implemented ClusterAD-Flight, which uses cluster based anomaly detection on pre-processed flight data parameters to identify flight level anomalies.

Typically techniques that are used to identify instantaneous anomalies are different than those used to identify flight-level anomalies. Supervised learning methods such as Inductive Monitoring System [18] (IMS) rely on a training set consisting of typical system behaviors which is compared with real-time data to detect anomalies. Each point is monitored standalone and therefore, the temporal aspect of anomalous sub-sequences is lost when identifying anomalies. Orca [19] is a technique that uses scalable k-nearest neighbor approach to detect anomalies in data with continuous and discrete features. Since each data point is a sample in time and treated as independent by the algorithm, Orca struggles to detect anomalies with temporal signatures. Amidan and Ferryman [20] have utilized Singular Value Decomposition (SVD) to identify instantaneous anomalies. They mapped the five seconds before and after each recorded data point and fit a linear regression model to it. The coefficients of the regression model were then used to create a mathematical signature for each recorded data point which was used to identify outliers. Mugtussidis [21] has used

Bayesian classification to distinguish between typical data points, that are present in the majority of flights, and unusual data points that can be only found in a few flights. Some of the methods rely on developing approximate models using flight data and detecting those flight records which deviate greatly from this model as outliers. Melnyk et al. [22] developed a vector auto regressive exogenous model to detect anomalies from flight data. Instantaneous anomalies are detected as residuals from the model. However, this technique involves identifying a model for each flight in the database and then calculating the residuals with respect to every other flight. Li et al. [23] developed ClusterAD-Data Sample, which is a technique leveraging a mixture of Gaussian models to identify probability of a sample being anomalous during take off, approach, and landing. This method also treats each data sample independently and uses additional context to identify whether a particular gaussian component is appropriate at a given time.

Various techniques for identifying instantaneous anomalies in time-series have been explored previously. Majority of these techniques focus on univariate time-series data by moving a sliding window through the time-series. For example, Keogh et al. [24] have used a sliding window based technique along with Symbolic Aggregate ApproXimation (SAX) representation of time series to identify unusual sub-sequences. SAX discretizes the continuous data stream into a word by using average values of the parameter in a given window. One of the important assumptions in some of these applications is that the time series are stationary i.e. their mean value does not change over time. This is not necessarily true for most parameters in flight data. In some applications where multivariate time series are considered, they are first converted into a univariate time series. For example Chandola [16] uses the technique of subspace monitoring to find the distance between successive windows (sub-spaces) using the principal angles of each subspace. Once this has been done, the multivariate time series is converted into a univariate series and a number of techniques are used. While the temporal aspect of anomalies is preserved in these techniques, data in each time series is compared only to the data from the same time series. This approach can result in loss of potential insights that can be gained from all other samples in the data set.

Outside of the aviation safety domain, a variety of anomaly detection techniques have been implemented in other domains. It has been used to generate anomaly detection algorithms from surveillance data for terminal airspace operations [25], Khalastchi et al. [26] have demonstrated online data-driven anomaly detection in autonomous robots using a sliding window with the Mahalanobis distance measure. Vespe et al. [27] have used unsupervised learning to understand maritime traffic representation and thereby understand anomalies as low-likelihood behaviors. Chakravarty et al. [28] have demonstrated anomaly detection and tracking in patrolling robots. In the driverless car domain work has been done on intrusion detection and cybersecurity and safety [29, 30], sensor anomaly detection and identification [31], abnormal vehicle event detection [32], etc. This paper aims to leverage the advances from these different domains in anomaly detection to apply to the problem of instantaneous anomaly detection in GA using energy-based metrics.

The work reported in this paper builds upon various previous efforts by the authors and collaborators. The genesis of this work was in the identification and survey of energy-based metrics for safety analysis in GA operations which has been laid out in Puranik et al. [33, 34]. For scenarios when an accurate estimation of the aircraft's performance (such as Thrust and Drag) is unavailable, it needs to be estimated using empirical models such as those developed by Harrison et al. [35] and Min et al. [36]. In order to accurately estimate some of the metrics surveyed for phases of flight such as approach, a reference or standard profile was required. While such reference profiles are common in commercial aviation operations, in GA no such profiles are readily available. Therefore, work outlined in Puranik et al. [37] used data-driven methods to define reference or standard profiles. Building upon this core work, techniques for detecting anomalies among thousands of flights obtained from heterogeneous GA operations were explored. Flight-level anomalies were examined in detail in Puranik et al. (using clustering) [38] and Puranik et al. (using a combination of clustering and one-class support vector machines) [39]. A comparative analysis of flight-level and instantaneous anomalies has been performed in previous work [40]. In addition to these quantitative techniques from flight data recorder data, work was also performed to understand qualitatively the nature of unstabilized approach in GA using accident data from the NTSB in Rao et al. [41]. The work reported in this paper builds upon the previous work by providing a detailed treatment of instantaneous anomalies using the identified energy-based metrics and raw flight data for multiple phases of flight as well as a comparison with exceedance detection approaches.

B. Gaps and Research Objective

There are various challenges in GA operations that preclude the use of some of the existing techniques in literature. First, the fleet of aircraft operated within GA is heterogeneous and even within the specific category of aircraft considered, the gross weight, rated engine power, and other aspects of the aircraft can vary considerably. In addition, some aircraft are possibly limited in terms of parameters recorded by their digital flight data recorders (if they have one installed) [40].

Therefore, metrics used in anomaly detection algorithms need to be those that correspond to safety margins and safe operations in GA and can be readily estimated from available recorded data. Therefore, using recorded flight data with quantitative aircraft performance models (such as those developed in Puranik et al. [42]) is a key enabler for safety analysis in GA. These models are used for evaluating a number of energy-based metrics, which provide a measurable quantification of the energy state of the aircraft and can be viewed as an objective currency to evaluate various safety-critical conditions across a large fleet of aircraft with minimal amount of recorded parameters [33, 34]. Energy metrics provide a unique combination of parsimony, generalizability, and relevance to safety that is appropriate for this application. Finally, the types of missions flown in GA can be quite different and non-uniform. Therefore, techniques that do not make assumptions about parameter variations in a phase of flight or are more broadly applicable across different phases of flight are desired. In most aviation applications, machine learning methods have been demonstrated on a fleet of similar aircraft (sometimes at the same airport). Therefore, it is unclear how well they generalize for different aircraft operating at different airports. Additionally, most approaches in literature are pure data-mining based in that they only utilize the recorded data and therefore do not use any additional domain expertise (such as energy metrics).

In addition to the challenges specific to GA identified in the previous paragraph, there are certain limitations of existing instantaneous anomaly detection techniques surveyed. In many approaches for identifying instantaneous anomalies, each point is monitored as a standalone independent sample and therefore, the temporal aspect of anomalous sub-sequences may be lost. Some approaches in natural language processing or machine translation such as recurrent networks with memory factor or forgetting factor (like the Long Short-Term Memory - LSTM [43]) explicitly try to account for temporal nature and context of incoming samples. However, these limitations are still observed in approaches like exceedance detection, where parameter values exceeding certain static thresholds are flagged as exceedances without necessarily considering their context. Moreover, in many anomaly detection applications, multivariate time series are converted into a univariate series or a high dimensional vector thereby causing some information to be lost. In some cases of monitoring, the time-series is only compared to reference thresholds or data from within the same time-series thus losing out on potential insights that can be gained from other flights.

The methodology developed in this paper aims to address some of the limitations of existing techniques to identify instantaneous anomalies and demonstrates one of the first applications in the GA domain. The following are identified as the main objectives of this paper:

1) Demonstrate the automatic identification of instantaneous anomalies in GA operations without a-priori input using energy-based metrics
2) Provide a continuous score – such as probability of being nominal – at every point in a flight record utilizing temporal information and correlation between parameters that might available in the flight data
3) Provide a comparative assessment against exceedance detection techniques to highlight benefits and limitations of developed method

III. Methodology

Unlike other applications of data mining or anomaly detection, aviation data is typically not labeled. There have been recent efforts to use non-linear dynamics models in lieu of actual accident flights to generate a database of failure-labeled trajectories in order to facilitate supervised learning [44]. However, in GA flight records, typically there is no knowledge a-priori as to which flight records (if any) are actually anomalous. Therefore, the anomaly detection algorithm needs to be semi-supervised or unsupervised. While an anomaly is a deviation from nominal or known behavior, the means of identification and context may differ. In the methodology presented in this paper, instantaneous anomalies are those instants that deviate from the nominal operations as identified by a clustering algorithm. The chief assumption in this type of application is that majority of the data contains nominal operations which is a reasonable assumption in this case because of the extremely low accident and incident rates per million flight hours in aviation (even GA) [1]. Noting these requirements, the data obtained from GA flights is analyzed in a general anomaly detection pipeline shown in Figure 1. Various components of this pipeline as utilized in this paper are described further.

A. Flight Data

In an anomaly detection problem it is important to understand the nature of the data captured (from flight data recorder in this case) as it has an impact on the techniques that can be used to analyze the data. Digital Flight Data Recorder (DFDR) data typically contains different channels that record discrete, continuous, and categorical data at a specified frequency (e.g., once per one second interval). Therefore, in analysis, the data obtained is a multi-variate time

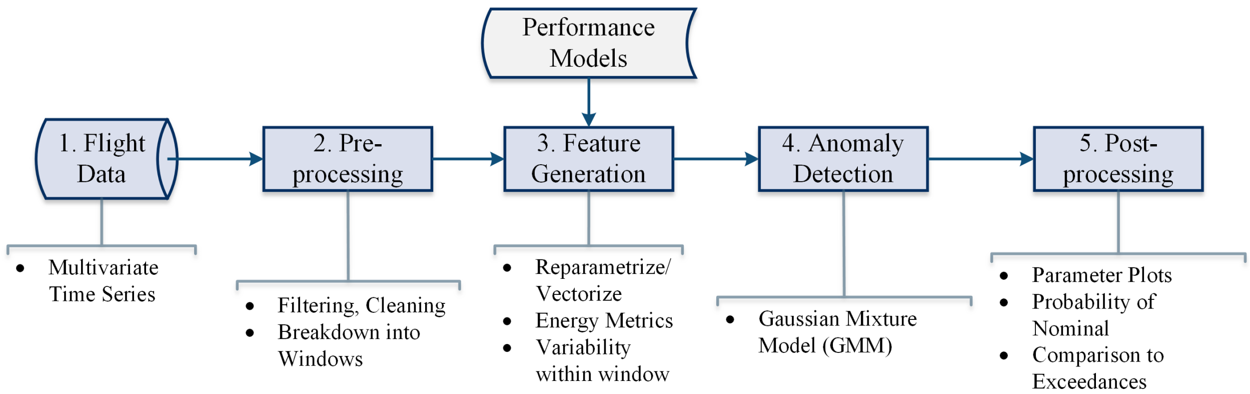


Fig. 1 Pipeline for anomaly detection used in this work

series for flight records, which are typically of different duration. The number of parameters in this multi-variate time series can be as low as 20 — 30 in GA operations to thousands in commercial operations [45, 46].

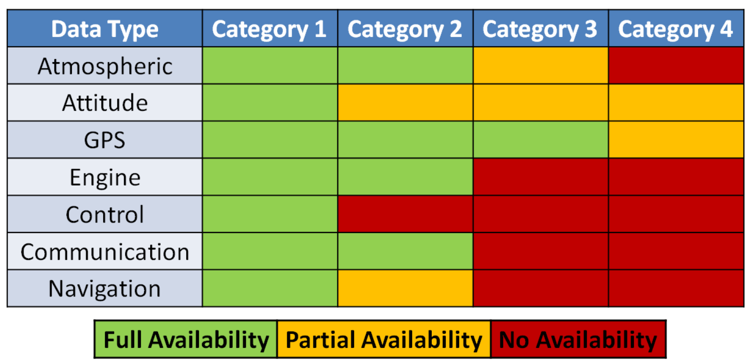


Fig. 2 Different categories of data availability in flight data recorders

The capabilities of flight data recorders can vary substantially; with GA typically lagging behind commercial aviation due to constraints on cost [10]. Figure 2 sorts the flight data recording capabilities into different categories based on the parameters from various systems that are typically recorded. The capabilities are sorted into four categories from one to four, with decreasing number of recorded parameters and fidelity for higher-numbered categories. The parameters are also divided into different types loosely based on systems in the aircraft. Note that this chart is intended to provide a notional example of data recording capabilities — specific recorders may provide more or less capability.

A brief description of the types of parameters in each system is provided here. Atmospheric data refers to data gathered from pitot tubes, barometers, thermometers, etc. It includes airspeed, wind speeds, pressure altitude, atmospheric temperature, etc. Attitude data refers to roll, pitch, yaw angles and their corresponding rates and accelerations. GPS data contains the latitude, longitude, altitude, and related rates. Engine data contains RPM, Exhaust Gas Temperatures (EGT), Cylinder Head Temperatures (CHT), oil temperature and pressure, fuel flow rates, fuel quantities, etc. Control data contains the deflection of flaps, elevator, aileron, rudder, etc. Communication data includes details about the communication status of the vehicle, such as the comm frequency. Finally, navigation data includes information on any waypoint guidance or autopilot features. In the context of Figure 2, full availability indicates all parameters from a particular system being available at a high fidelity, partial availability indicates that the some of the parameters of that system might not be recorded or are recorded at a lower fidelity or sampling rate (for example: data is collected from a personal electronic device (PED) as opposed to a Garmin G1000 glass cockpit). No availability indicates that the parameters from that system are not recorded in any form.

In the development of any new methods of safety analysis for GA domain, the capabilities of flight data recording devices is one of the most important limitations. The spectrum of possible parameters being recorded in GA is fairly wide and heterogeneous. There has been recent research into investigating the use of low-cost personal electronic devices for data collection and analysis [47, 48]. While the typical GA aircraft lacks a sophisticated avionics system, the use of PEDs such as tablet computers by GA pilots has become increasingly popular. These can provide access to the Global Positioning System (GPS) data of the aircraft which includes latitude, longitude, and altitude data alongside heading and ground speed. If in addition the pilot utilizes an external attitude and heading reference system (AHRS), such as the commercially available Stratus [49], additional flight data can be collected by the PED. Similarly, at the higher end of the spectrum are glass cockpit systems such as the Garmin G1000 [50] which are able to log many more parameters accurately.

B. Pre-processing

Figure 2 in the previous subsection provided a general idea of the spectrum of possible parameters that can be recorded but did not specifically indicate where GA aircraft typically lie. A brief survey of parameters recorded in GA indicates that the actual parameters that can be recorded by typical GA flight data recorder is rarely at the sophistication of category 1. Data recorded from GA operations considered in this work fall under categories 2,3,4. The higher end of the GA spectrum such as G1000 can be considered category 2, whereas more common among GA airplanes would be category 3 or 4 recording capability. It is also noted that some GA aircraft may lack any data recording capability. In the present work data collected for the G1000 is used for developing the methodology, however, metrics that are more widely usable are leveraged in order to make the method more generically applicable. Energy-based metrics are utilized in this work as they provide a measurable quantification of the energy state of the aircraft and can be viewed as an objective currency to evaluate various safety-critical conditions across a large fleet of aircraft with minimal amount of recorded parameters. Table 4 in the appendix provides a brief summary of utilized energy metrics and the data required for their computation.

The raw data obtained from flights is pre-processed to obtain data suitable for anomaly detection. Sensor noise that may be present in the original data is smoothed using a moving average filter. Previous work [37] is leveraged on ideal parameters to use for smoothing GA flight data. Typical methods for identification of instantaneous anomalies treat each data sample independently. This results in the loss of temporal nature of data obtained from flight records. It is of interest to capture the variation of key metrics before and after the time stamp under consideration. Therefore, in this methodology sliding window techniques are used for pre-processing flight data. This technique has been used previously on univariate time series in other domains [24, 51]. It involves sliding a window of a particular duration across the length of the time series (see Figure 3 for notional example).

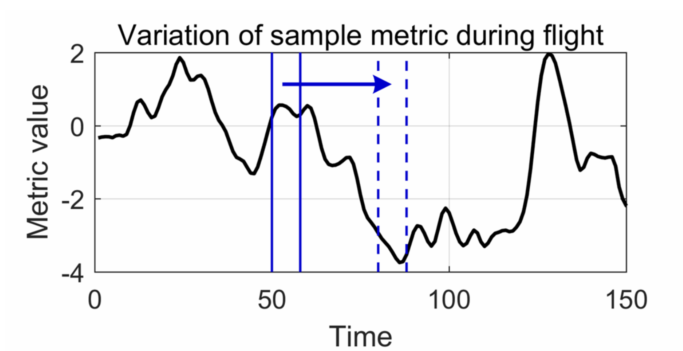


Fig. 3 Notional depiction of sliding window across a metric

For each time stamp, the data contained within the entire window is utilized rather than just the time stamp under consideration. There are various factors that can affect the choice of length of the window such as total duration of the time series, typical response time of the system, computational concerns etc. Due to all these factors, there is no consensus in the literature as to the appropriate length of the window [51]. A small window size will result in inadequate capture of temporal aspects and a large window size results in over-smoothing. The definition of how small an interval of time qualifies as instantaneous is not always clear in literature. For example, there is a chance that gusty/turbulent conditions can result in a lot of flags for instantaneous anomalies. An airplane caught in a thermal/downdraft/wind-shear, etc. may exhibit rapid changes in airspeed or vertical speed. Instantaneous anomalies should be able to identify these conditions while being insensitive to random (unavoidable) noise in sensor measurements. Incorporating the variability of a metric in a window helps avoid some of these false flags due to noise in the data set. In this work, five seconds (two seconds before and after the current point except at the ends of the series) is used as the window length. This number is obtained based on computational time typical response times of GA aircraft and sensor noise in GA data [52, 53]. An important distinction between existing window-based techniques in other domains and the application in this paper is that window-based anomaly detection typically tries to find anomalous windows in a time series based on the data in that time series itself. On the other hand, the current method aims to also leverage the additional information available from the rest of the data set.

C. Feature Generation

One of the most important steps in the anomaly detection framework is the generation of *features* for the algorithm. Usually, these features are derived from the manipulation of the raw data collected. This can be as simple as using the raw data directly, to calculating new metrics using a combination of the recorded data and external information. The performance of the algorithm in identifying appropriate anomalies can depend heavily on the types of features used [16]. In previous work (Puranik et al. [33, 34]), it was demonstrated that energy-based metrics such as those quantifying the energy state, rates of change of energy, and their margins and deviations are appropriate metrics that correspond to safety margins and safe operations in GA and can be readily estimated from recorded data. Therefore, in the present work, energy-based metrics are utilized as features for identification of instantaneous anomalies.

In a window-based techniques, data from each window is used to augment the flight data (or energy metrics) recorded at time stamp. In order to obtain features of each window, the mean values of energy metrics along with their variability within the window is calculated. This allows the temporal aspect of instantaneous anomalies to be captured while detecting anomalies. Specifically, the range of the metric values (maximum value minus minimum value) within the window is used as a measure of its variability. Thus, for example, for each time stamp in the flight data if there are k energy metrics being used, then the feature vector for that window will contain $2 \times k$ dimensions (the mean value of each metric (m) and its variability(v) within the window). Equation 1 shows the feature vector for each time stamp which is then subsequently used in anomaly detection.

$$\mathbf{f} = \left[\underbrace{m^{(1)}, v^{(1)}}_{\text{Metric 1}}, \underbrace{m^{(2)}, v^{(2)}}_{\text{Metric 2}}, \ldots, \underbrace{m^{(k)}, v^{(k)}}_{\text{Metric k}} \right] \tag{1}$$

Thus, a feature vector is generated by concatenating the mean values of each energy metric at that time stamp along with its variability within the sliding window for each point in the time series and subsequently used for instantaneous anomaly detection.

D. Anomaly Detection

The next step is to use the features generated for each point within an anomaly detection algorithm. It is noted previously that most anomaly detection techniques for instantaneous anomalies did not deal with multivariate data explicitly and converted it into univariate data prior to analysis. However, it is desired that algorithms deal with multivariate data directly be used. Therefore, a Gaussian Mixture Model (GMM) is used for clustering and anomaly detection. The GMM can cluster normal operations together and help identify anomalies along with their probability based on the data set. One of the main advantages of using a GMM is that it operates directly on the multivariate series and does not transform it into a univariate series. The other advantage is that GMM allows multiple standard operations to simultaneously exist. This is very important for GA applications where heterogeneity in operations can cause false positives in machine learning models. Additionally, since the Gaussian Mixture Model is able to cluster and provide a probability score for each point which is a continuous function over the flight record, sudden gusts/noise will not be able to perturb this measure as easily.

A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. Each component in the mixture is a type of standard observation or behavior of the system (example, one component for each phase of flight). The number of components, k, in the GMM determines number of sub-populations or clusters. The relation between features is captured in the form of a covariance matrix Σ. If each member of the population (in this case each feature vector) is an m-dimensional vector, then the GMM with k components and a covariance matrix Σ is given by:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{k} w_i g(\mathbf{x}|\mu_i, \Sigma_i) \tag{2}$$

where

$$g(\mathbf{x}|\mu_i, \Sigma_i) \tag{3}$$

indicates each of the components of the mixture model which is a multivariate gaussian model and w_i indicates the weighing of the component. The trained GMM is completely defined by the three parameters (w_i, μ_i, Σ_i) and k – the number of components. The parameters of the GMM are typically obtained via an Expectation-Maximization (EM) algorithm using the data set available [54]. This technique is utilized in the work presented in this paper. However, there are a few other important decisions that need to be made regarding the nature of the model before the EM algorithm can be used. These are the nature of covariance matrix (full or diagonal), whether parameters among gaussian components are shared or not, and the number of components k. Due to the large computational cost of full covariance matrices, diagonal covariance matrix is used and the parameters are not shared among different gaussian components. Similar decisions have previously been used by Li et al. [23] on application in commercial operations. Finally, the number of components can be set based on prior knowledge or it can be obtained using statistical metrics. Identifying the appropriate number of components using sensitivity analysis can be accomplished using different internal clustering criteria which use the distance between different points (usually Euclidean distance) to provide a measure of quality. This measure is based on various factors like total dispersion, within-cluster similarity, and between-cluster dissimilarity. With internal clustering criteria rugged topography might be more important in decision-making than the magnitude of the criterion itself. The choice of which criterion to use is dependent on the nature of the data set available and its properties. In the current work, different indices are explored and the one with the clearest landscape is chosen for making the decision on the number of clusters. This criterion is the Calinski-Harabascz index (C-H index). C-H index is most suitable for cases where clusters are more or less spherical and close to normally distributed.

The advantage of using GMM for clustering is that it can provide statistical inferences about the underlying distributions. Therefore, once the required GMM has been trained using the existing data, it can be used to detect outliers or anomalies among the flights. Using the values of the parameters obtained for the GMM (w_i, μ_i, Σ_i) the posterior probabilities of any component p for an observation $\mathbf{x}$ can then be calculated as:

$$p(\mathbf{x} \in p) = g(\mathbf{x}|\mu_p, \Sigma_p) \tag{4}$$

The estimated probability density function for each observation is then obtained as a sum over all components of the component density at that observation times the component probability.

E. Post-processing

Using the estimated probability density function of each time stamp, a profile of the probability density over the entire duration of the flight as shown in Figure 4 can then be constructed. Using appropriate thresholds for the probability enables identification of anomalous sub-sequences or instantaneous anomalies.

Appropriate thresholds in this context refers to the trade-off between missed detection (high type II errors) and excessive analyst workload. The recommendations on thresholds (such as the 0.05% or 0.1%) are obtained using the available data and thus could be biased by the data available. In order to limit potential bias from the available data, it is recommended that data be obtained from a variety of aircraft, airports, and operators. However, that may not always be possible and is thus a limitation of the method. This threshold can thus be varied to obtain different numbers of instantaneous anomalies. The safety analyst can then decide this threshold based on the trade-off between number of anomalies and missed detection.

The type I errors or false positives get managed through the setting of the threshold for anomaly detection. Since these thresholds are not arbitrarily selected by obtained by using values within the data itself (0.05 or 0.1 percentile), they get managed at the source. Once instantaneous anomalies are identified, plots of variation of flight parameters and energy metrics can be used to visualize and understand the reason for identification of this anomaly. This flexibility enables the analyst to focus attention on a limited number of important anomalies as opposed to a large untenable data set. The identified anomalies can also be compared against traditional exceedance events such as those defined in Table 3 in the appendix.

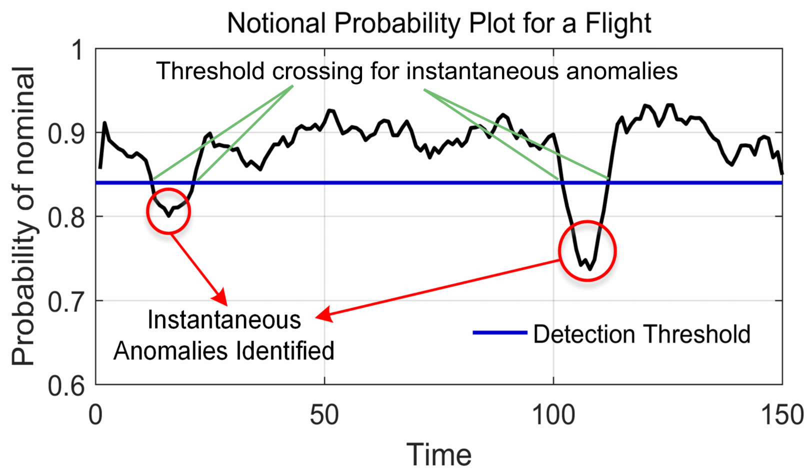


Fig. 4 Notional depiction of probability density at each point during a flight record and the detection threshold for identifying instantaneous anomalies

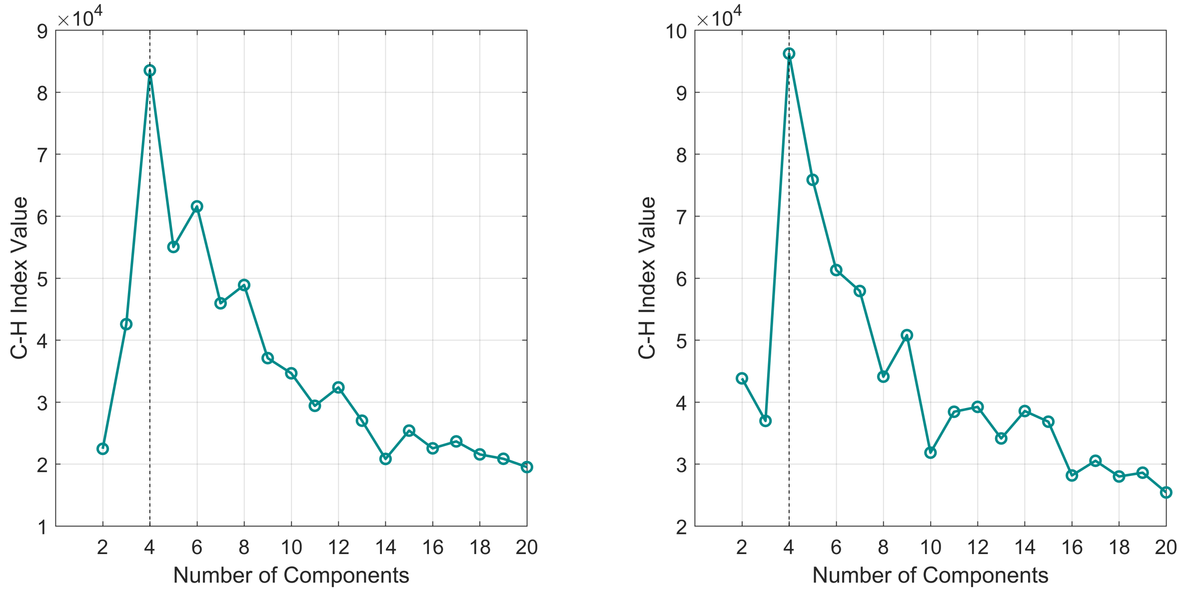
IV. Results

In this section, approximately three thousand flight data records obtained from two different GA flight schools collected during routine operations are utilized. The data are from two different types of aircraft (Cessna 172S and Piper Archer) operated at multiple GA airports. There are multiple aircraft of each type (different tail numbers) within the data-set. The data is processed as described earlier in the anomaly detection pipeline (Figure 1). While instantaneous anomalies can be identified for all phases of flight, in this paper, the two most safety-critical phases of flight are considered – take off and approach and landing. This section contains three parts: First, the selection of model parameters using sensitivity analysis for this particular data set is obtained. Next, specific examples of anomalies obtained using the selected parameters in the two phases of flight are shown along with accompanying explanation and possible causes. Finally a comparison of the developed anomaly detection method with traditional exceedance detection techniques is presented highlighting its benefits and limitations.

A. Algorithm Parameter Selection

The first step in instantaneous anomaly detection is identifying the approximate number of components that will be used in the Gaussian Mixture Model. As noted earlier, this number is typically obtained using sensitivity analysis based on some information-theoretic indices and depends upon the data set that is used. Therefore, unless a given data set contains all possible variations that can affect operations (weather, airport, aircraft type, number of passengers, etc.) this step will have to be repeated for a new data set in order to confirm the ideal number of components. For this step, the GMM is trained with steadily increasing number of components and the C-H index is measured for each trained model. The model with components that gives the highest C-H index is the one displaying the best internal clustering structure and is chosen for this application. This sensitivity analysis is performed for both the take-off and approach-and-landing phases. The results obtained are shown in Figure 5.

Based on this sensitivity analysis, the number of components chosen for the GMM for both takeoff and approach-and-landing cases is **four** as this number has the highest value of the C-H index. While each component of the GMM typically represents a standard behavior of the system, an initial assessment of the four components in this paper did not present any distinctive behavior. This is due to the way the feature vectors are set up. The mean and variance of each Gaussian component and the mixing probabilities for the trained GMM with four components are then used in further analysis. These parameters are then utilized to calculate the posterior probability density of being nominal at each point in the appropriate phase of flight using equation 4. Once a posterior probability is obtained for each point, different thresholds can be set of detecting instantaneous anomalies. Instantaneous anomalies are identified as those points for which the probability density falls below a selected threshold. The detection threshold for anomalies can be varied depending on the trade-off between type-I (false positive) and type-II (false negative) errors. Since the data is obtained from routine operations, there is no ‘ground truth’ available to compare anomalies, however, it is expected that the number of anomalies (if any) will be a small fraction of the total data as all the flights landed safely.



(a) Take-off

(b) Approach and Landing

Fig. 5 C-H index sensitivity with increasing number of components for different phases of flight

B. Instantaneous Anomaly Examples

The next two sub-sections demonstrate the probability plots, energy metric plots, and parameter variations for a few instantaneous anomalies identified in the take-off and approach and landing phases in the data set. Plotting the variation of energy metrics and parameters during the phase of flight can help understand what part of the flight contained the instantaneous anomaly and compare how it performed relative to the entire data set. For post-processing the same style of percentile plots is presented for the instantaneous probability density, energy metrics, and raw flight parameters. In each figure, the dark grey bands represent 50^{th} percentile of all flight records and the light grey bands represent the 95^{th} percentile of all flight records. The solid black line represents the specific flight data under consideration. Overlaying it on the percentile bands enables comparative assessment of how a particular flight performed at each instant compared to every other flight in the data set. The distance remaining to the displaced runway threshold displayed on the x-axis for approach and landing and the time is displayed on the x-axis for take-off phase. For each phase a subset of interesting energy metrics and parameters is visualized. Two detection thresholds are displayed in the probability plot for the purpose of illustration and these correspond to 0.05% of all points and 0.1% of all points in that phase of flight. This implies that when the instantaneous probability density falls below these thresholds, it is lower than the probability values of most points in the data set and is therefore a very rare occurrence.

1. Approach and Landing: Flight with one instantaneous anomalous window

Figure 6 shows the variation of the probability density (on logarithmic scale) as a function of the distance remaining to the runway threshold for the flight record containing the instantaneous anomaly. The higher the probability, the more normal or nominal that point of the flight is compared to all other flights in a similar situation. The variation of probability density indicates a flight with nominal variation for most parts of the flight other than a small window when the distance remaining to the landing runway threshold is approximately two nautical miles. The anomaly is identified under both detection thresholds mentioned earlier. Beyond the indicated region of anomaly, the flight data record recovers its nominal behavior through to the end of the flight.

Figure 7 represents the variation of the energy metrics for the flight data record. The figure clearly indicates fluctuations in several metrics during the instantaneous anomaly, notably, potential, kinetic energies, and the kinetic energy rate. The profiles indicate a quick recovery to the nominal variations for these and other metrics after the instantaneous anomalous window. The raw parameters for this flight data record are visualized in a similar manner in Figure 8. It is evident that the RPM suddenly drops to well below its previous value and in order to maintain the altitude

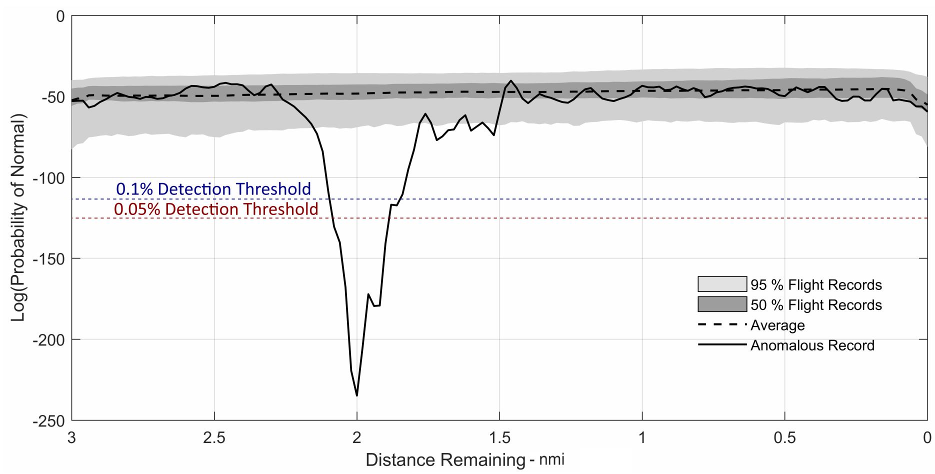


Fig. 6 Probability density at each point during approach and landing for a flight record with instantaneous anomaly and the detection thresholds

profile, the aircraft is pitched up. Vertical speed also becomes greater than zero in this phase which is highly unusual for approach and landing. The high kinetic energy and true airspeed all the way through to the end of approach indicate that the flight landed very fast and may have been a landing with no usage of flaps. For this particular flight record, flap position is recorded in the data set and this hypothesis is confirmed as the flap position remains zero throughout the approach and landing. Upon examining the exceedances for this flight record, it is observed that of the fourteen points in the instantaneous anomaly, there are **two** level-1 exceedances and **nine** level-2 exceedances. Therefore, it is worth investigating the actual cause of this instantaneous anomaly.

It is worth noting that while there is a precipitous drop in the instantaneous probability, not all energy metrics or raw parameters vary outside their nominal bounds in this region. In fact, the drop is caused by a specific combination of energy metrics varying outside their nominal bounds over the period of the drop in probability density. This is due to the fact that the method accounts for the multi-variate nature of the data while evaluating the probability score.

2. Take-off: Flight with one instantaneous anomalous window

Similar to the previous subsection, a flight with instantaneous anomalies in the take-off phase is examined here. Figure 9 shows the variation of probability density at each point during the take-off phase for the flight record under consideration. It is observed that the instantaneous anomaly occurs in the initial part of the take-off phase (approximately 40 seconds after take-off) where there is a large drop in probability density.

Exploring the variation of energy metrics during this flight in Figure 10 it is observed that during the anomalous window, the potential and kinetic energy rates display fluctuations out of nominal bounds as does the potential flight path angle. The potential energy rises faster than for nominal operations in the initial part which is being sought to be corrected during which the instantaneous anomaly occurs. The variation of raw flight parameters shown in Figure 11 presents a similar picture. The altitude near the instantaneous anomaly is leveled off by rapid reduction in pitch attitude to gain airspeed but it results in a negative vertical speed during the take-off phase which is considered highly anomalous by the clustering algorithm.

Similar to the approach and landing case, only a subset of energy metrics and parameters vary outside the nominal bounds despite the drop in probability density indicating that the multi-variate nature of the clustering is at work. Furthermore, the abnormal variations during the anomalous window are more apparent because the algorithm uses data from all other flights in the same situations. As will be elaborated in later sections, using exceedance detection results in a large number of false positives due to heuristic definitions.

3. Observations

The instantaneous probability density provides a single metric to be monitored that takes into account multiple parameters, features, and their correlations as well as data from all flights in the data set. Thus, it reduces the need to monitor several parameters at the same time which can be cumbersome. The points in the flights with lower probability

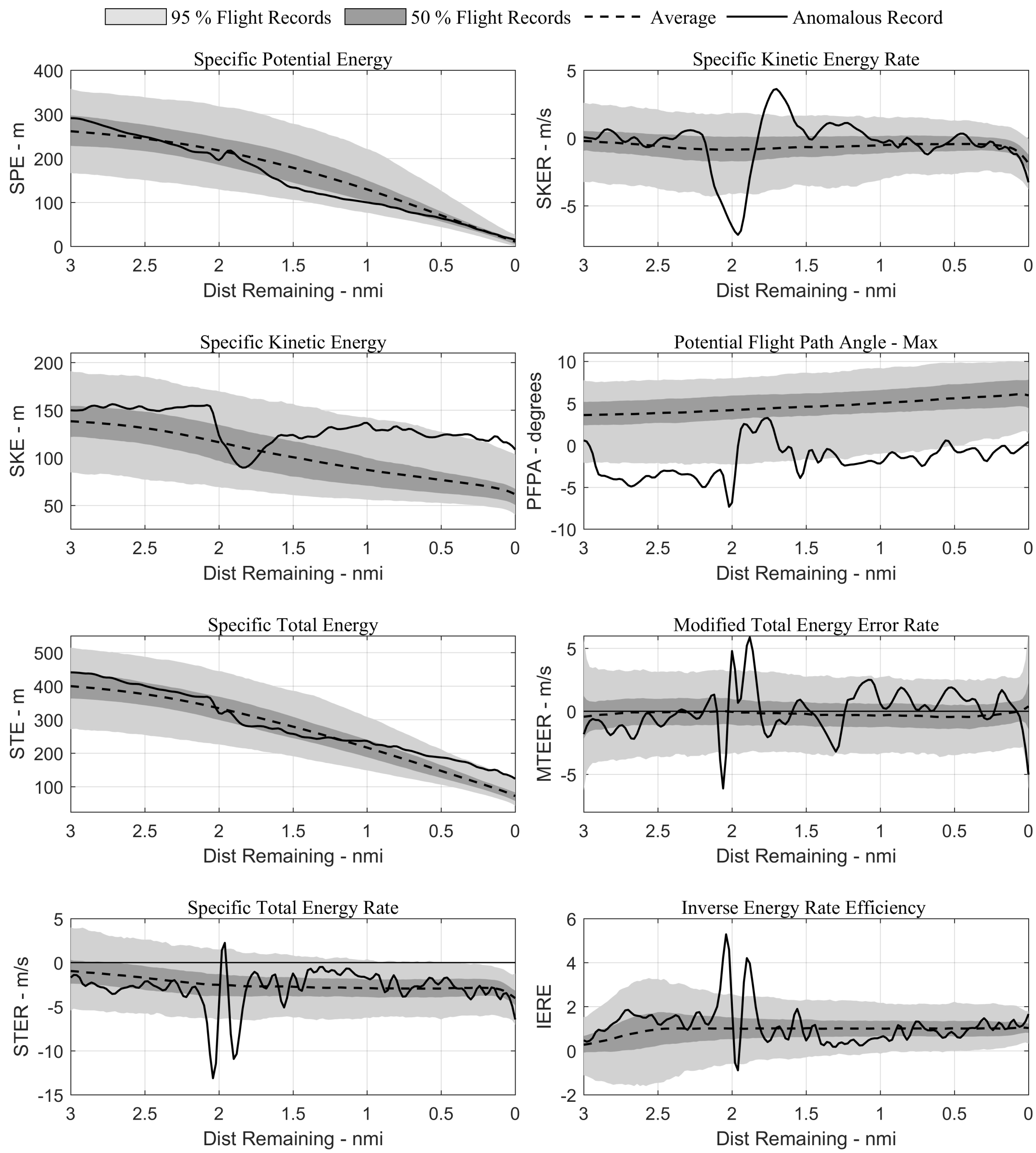


Fig. 7 Variation of energy metrics during approach and landing for a flight record with instantaneous anomalies

densities are always associated with abnormal parameter and energy state variations which can be direct precursors to accidents [55]. The flexibility of the detection threshold provides the much-needed control between type-I and type-II errors in detection.

While clear deviations are visible for both the raw parameters and the energy-based metrics in the identified anomalies, only the energy-based metrics as opposed to the full parameter set have been used for anomaly detection.

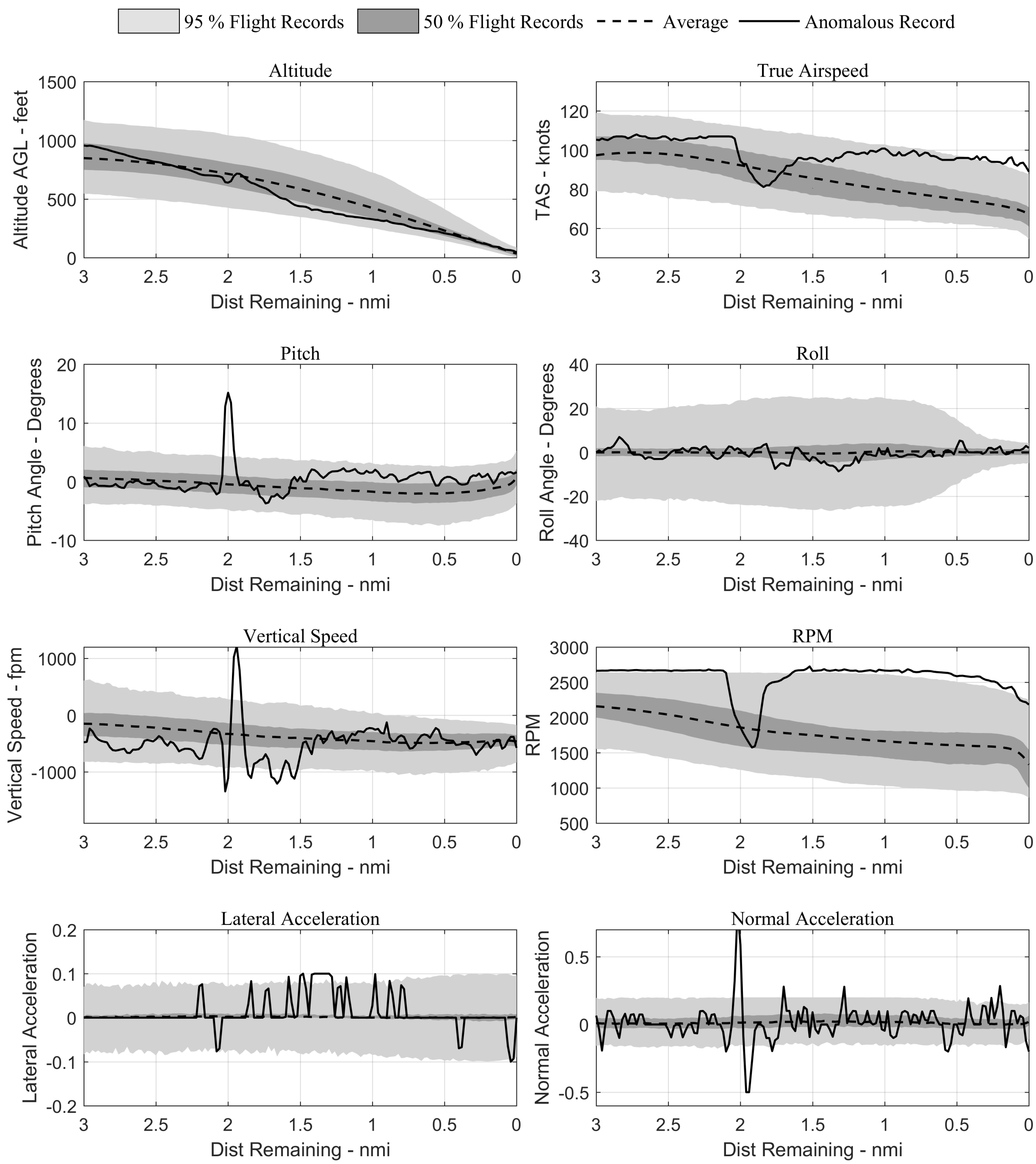


Fig. 8 Variation of flight parameters during approach and landing for a flight record with instantaneous anomalies

Thus, the method illustrates identification of the instantaneous anomalies in both phases of flight using a parsimonious set of energy-based metrics. Therefore, if energy-based metrics are used instead of raw parameters, the method can be applied more widely; even for aircraft that do not have a sophisticated data acquisition system. The minimum set of parameters thus required for the energy-based metric implementation in this paper is outlined in the appendix (Table 5).

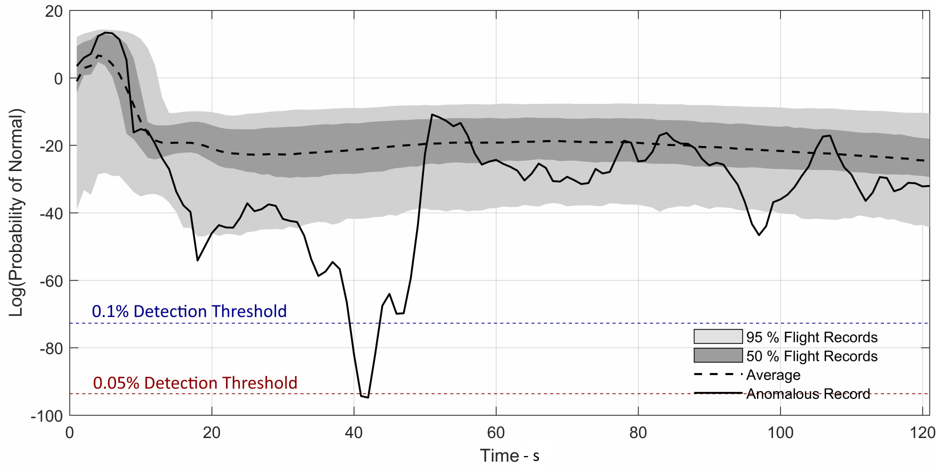


Fig. 9 Probability density at each point during take-off for a flight record with instantaneous anomaly and the detection thresholds

This set corresponds to the metrics from Table 4 that can be estimated using flight data and a performance model (last column of table). In the event of performance models being unavailable, this can be scaled back to the energy-based metrics obtained from pure flight data which only requires four parameters - altitude, true airspeed, vertical speed, and flight path angle.

Comparison between the results of the approach and landing and take-off phases reveals that there is a higher homogeneity in the initial phases of the take-off phase with high instantaneous probability densities compared to approach and landing. One of the reasons for this is that the initial part of the take-off phase has much higher homogeneity than any others in the data set. This is attributed to the fact that all take-off phases in the initial part are characterized by high power, increasing speeds (and kinetic energy), altitude gain (positive potential energy rate), etc. These similar conditions cause the cluster to be tightly knit and subsequent high probabilities. This is also evident in the absolute values of the probability densities for the two phases of flight.

C. Comparison to Exceedance Detection

Exceedance detection is the most common method of analysis using FOQA data currently in existence. An exceedance is the deviation of a single parameter beyond an established threshold. An event is defined by one or more parameter exceedances that take place concurrently over a specified period of time. Exceedance detection is designed to provide answers to questions that the domain experts have thought to ask. There can be a lack of traceability in event definitions and a need to continuously fine-tune safety events. Moreover, events are defined in limited number of dimensions, which usually correspond to the most critical parameter(s) and are not well established for improbable conditions. Greater fleet heterogeneity and types of missions in GA can increase chances of missed detection or false alerts without fine-tuning.

Table 1 Summary of exceedance events in approach and landing phases in the current data set

Total types of exceedances considered (From Table 3)	25
Proportion of Flights with at least one level-1 exceedance	92.5 %
Proportion of Flights with at least one level-2 exceedance	76.9 %
Average of number of level-1 exceedances per flight	11.55
Average of number of level-2 exceedances per flight	16.22

While this technique has these limitations, it is nevertheless of value to compare the flight data records with anomalies against already defined exceedances or events. This can provide insight into the ability of anomaly detection using energy

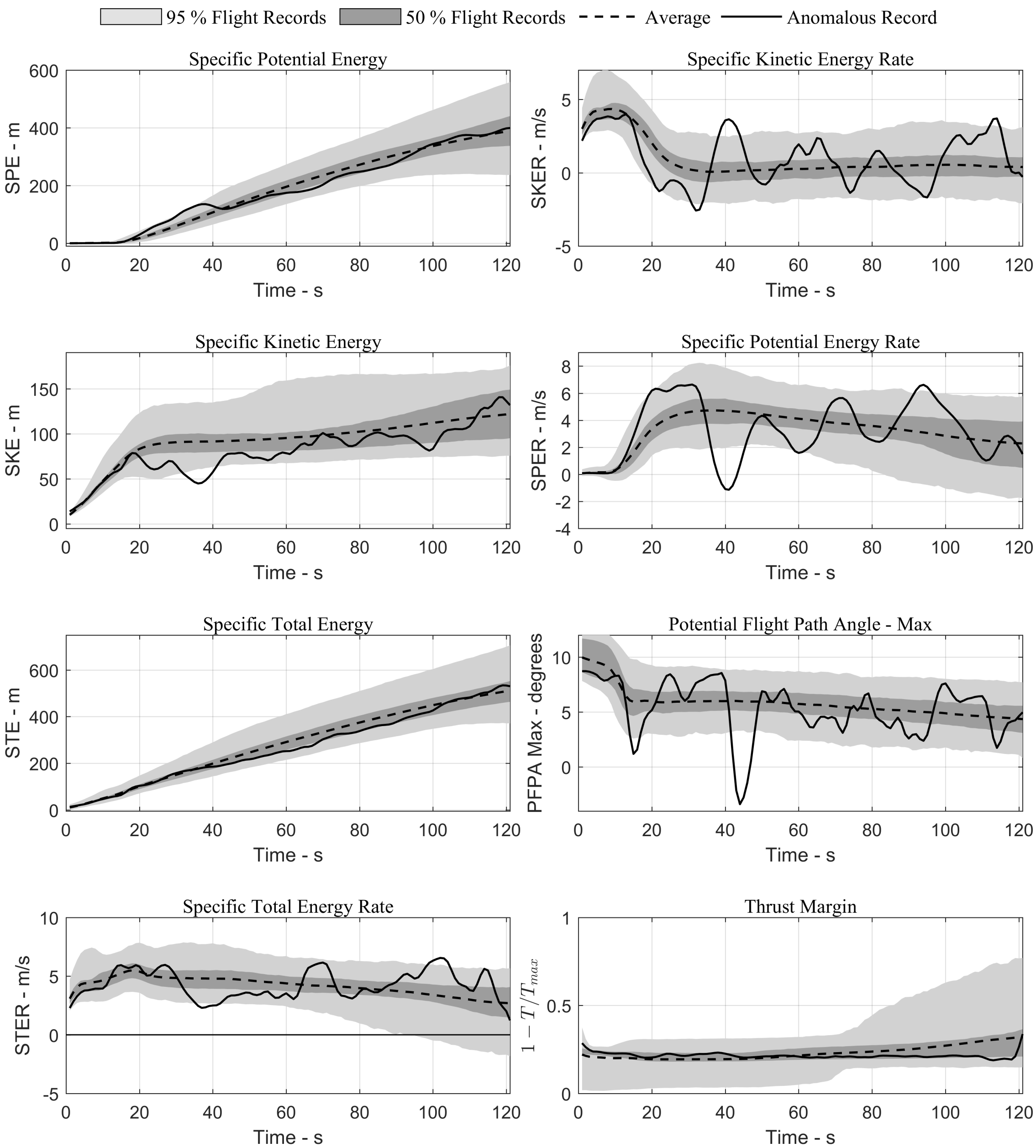


Fig. 10 Variation of energy metrics during take-off for a flight record with instantaneous anomaly

metrics to identify some generally accepted exceedance events. In the present section, comparison with exceedance detection is only provided for approach and landing phase as take-off phase has fewer documented exceedances that can be easily computed as noted in [56]. It is important to understand why exceedance detection is limited in its applicability in addition to the reason than that event definitions might not be transferable between different aircraft. For that purpose, all the flights in the data set are examined for exceedances as defined in Higgins et al. [56]. Each point in the flight

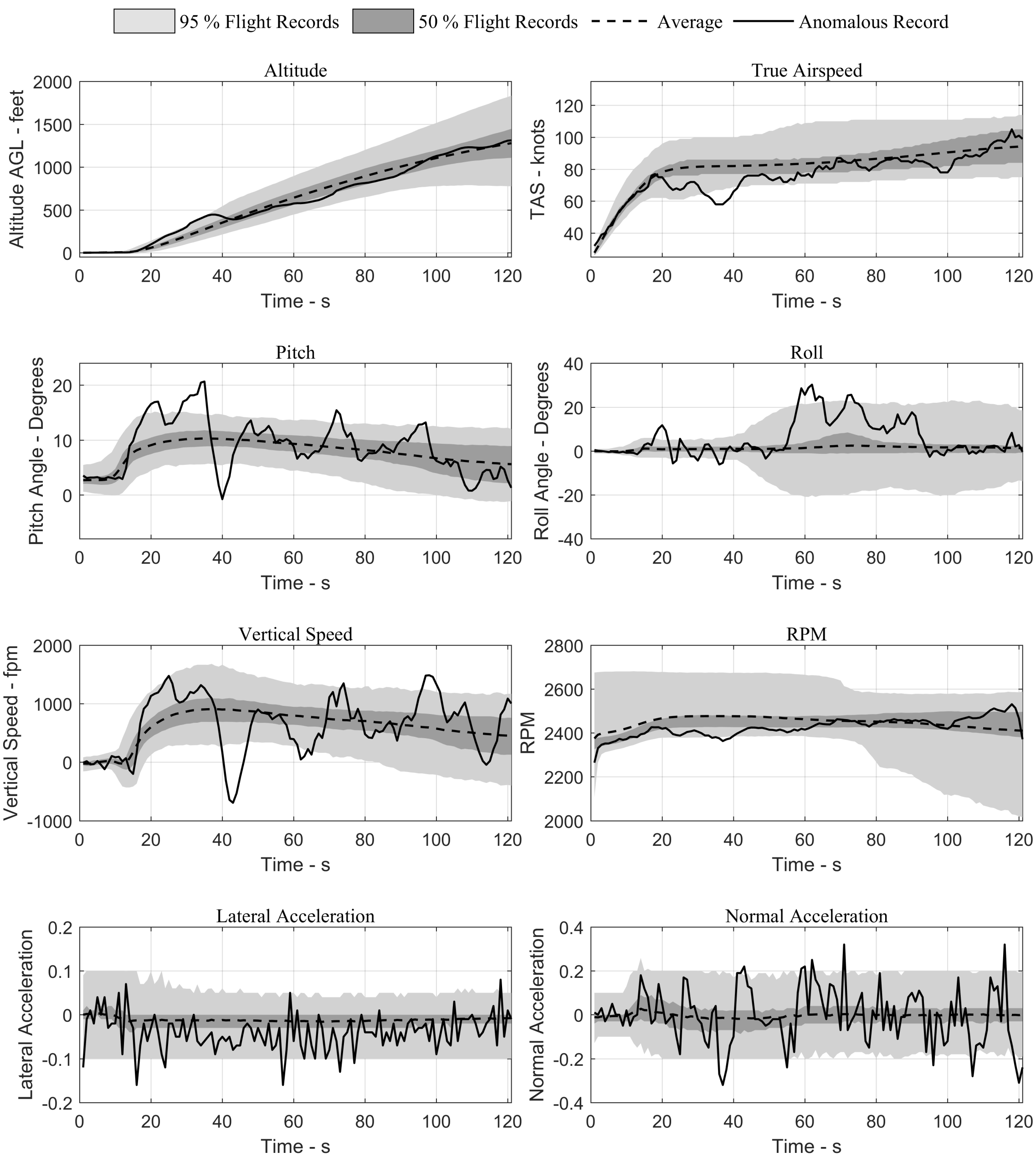


Fig. 11 Variation of flight parameters during take-off for a flight record with instantaneous anomaly

data record is examined as a standalone point to identify any level-1 or level-2 exceedance events that are occurring. The summary and some statistics for the total number of events for the approach and landing phase is presented in Table 1. The main reason exceedances from [56] are used is that it is currently one of the few publicly available set of exceedances in the GA domain.

As is evident from the table, a high number of flights from the current data set have at least one level-1 exceedance

(≈ 92.5 %) or at one level-2 exceedance (≈ 76.9 %). This indicates that based on the exceedances defined in Table 3 a high number of flights have at least one exceedance event. Despite the flight data being collected from routine operations and training flights, a high number of flights returning with exceedance events is unexpected. The relationship between instantaneous anomalies and exceedances can be explored using the data set. The average instantaneous anomaly probability for different types of points in the entire data set is obtained as shown in Table 2. It is observed that the average log probability of level-2 exceedance points (–51.08) is lower than level-1 exceedance points (–49.17), which is further lower than points with no exceedances (–46.36).

Table 2 Average instantaneous probabilities for different types of points in flights

Type of Point	Average Instantaneous Probability (logarithm)[a]
Points with no exceedance	– 46.36
Points with level-1 exceedance	– 49.17
Points with level-2 exceedance	– 51.08

[a]More negative value indicates more anomalous points

These results indicate that there is an overall negative correlation between the instantaneous probability score and the exceedance level. Points with lower probability score have higher exceedance levels, whereas points with higher probability score have lower exceedance levels. Those points with no exceedances typically have better instantaneous anomaly score than those with either type of exceedances. However, this correlation is of limited use. Figure 12 shows the distribution of instantaneous anomaly score for level-1 exceedance points, level-2 exceedance points, and those points with no exceedances along with the median for each distribution. Despite the slight shift in the distribution for exceedance points versus non-exceedance points, there is still a significant overlap between all three distributions. The typical thresholds for different levels of instantaneous anomaly detection are also shown in Figure 12. This leads to the following observation – a point with an exceedance may not necessarily be an instantaneous anomaly as seen from the thresholds being on the extreme left of the figure.

However, upon examination of each point with an instantaneous anomaly at the two different thresholds, it is observed that 100% of the instantaneous anomalous points at 0.05% anomaly threshold and 98% of the instantaneous anomalous points at the 0.1% anomaly threshold contain at least one exceedance.[†]

The main appeal of the anomaly detection approach is its usability in a variety of situations, with different aircraft, with multiple types of data recorders, etc. Anomaly detection is also aimed at providing a smaller, focused list of flights/regions to examine rather than the large set of points that might be obtained from exceedance detection. Thus, anomaly detection as it is undertaken in this paper is able to provide incisive insights into the instantaneous anomalous behavior of flights and is more usable than the existing benchmark technique of exceedance detection.

V. Conclusion

In this paper, a novel method for identifying instantaneous anomalies in GA flight data is demonstrated. This method utilized energy-based metrics as features in an anomaly detection framework. A sliding-window based pre-processing technique is used to ensure temporal aspects of features are captured. A mixture of gaussian models is trained using the available flight data for cluster analysis and outlier detection. Multivariate series are explicitly treated in the model and it also allows for multiple standard operations which addresses some of the limitations of previous methods. The proposed method is demonstrated on over three thousand flights collected from two representative single-engine naturally aspirated GA aircraft operating at two different airports. Examples of an instantaneous anomalies identified by the methodology in the take-off and approach and landing phase are presented along with instantaneous probability density plots and possible causes. The use of energy-based metrics ensures that the methods developed are generically applicable across

[†]It is noted that, for some of the flight records, it is not possible to evaluate the exceedance events as they did not have all the required parameters recorded. Therefore, in Figure 12, only those flight records for which it is possible to calculate exceedance events (approximately 70% of the total three thousand flights) are included.

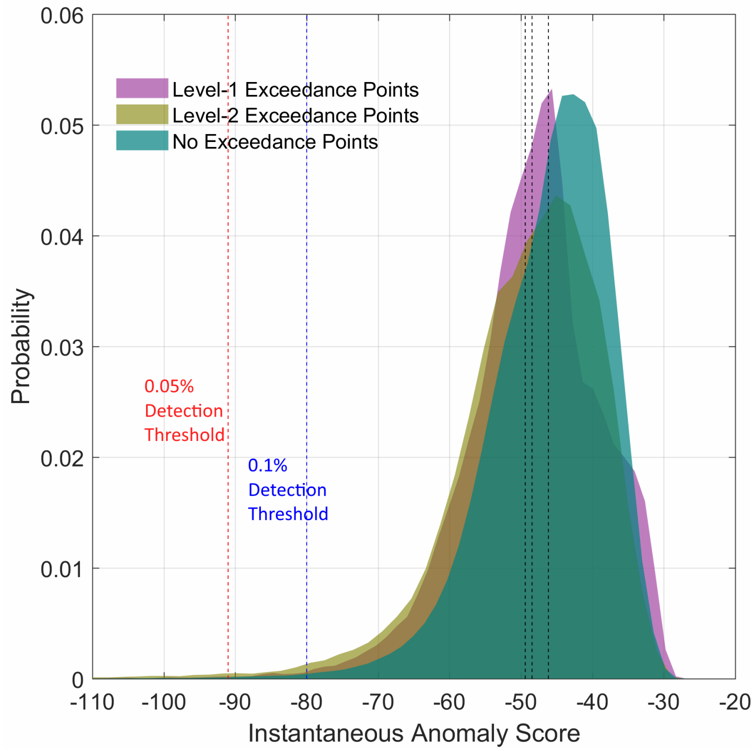


Fig. 12 Distribution of probabilities of level-1, level-2, and no-exceedance points

multiple aircraft with varying flight data recording capabilities. Finally, a comparison of anomaly detection with existing technique of exceedance detection is performed to highlight the advantage of the data-driven approach.

The main advantage of using these methods is that the expert review process is cut down due to the specific anomalous windows identified. The method demonstrates a uniform process for all phases of flight rather than different event definitions for each phase and aircraft. The instantaneous probability density provides a single metric to be monitored that takes into account multiple parameters, features, and their correlations as well as data from all flights in the data set. Thus, it reduces the need to monitor several parameters at the same time which can be cumbersome. It is acknowledged that while the identified anomalies exhibit deviations from nominal operations in multiple metrics, their relative significance from a safety standpoint would still subject matter expert review. The comparison of this type of traditional machine learning technique against modern deep learning techniques is a potential avenue of future work.

VI. Acknowledgments

This work was sponsored by the Federal Aviation Administration under Project No. 5 of the FAA Center of Excellence for General Aviation, PEGASAS (Partnership to Enhance General Aviation Safety Accessibility and Sustainability), via grant No. 12-C-GA-GIT-018 to the Georgia Institute of Technology. The authors are grateful for the support of project partners. The information in this research does not constitute FAA Flight Standards or FAA Aircraft Certification policy.

Appendix

A. List of Exceedance Events

Table 3 Sample exceedances set during approach and landing for a Cessna 172 aircraft (Higgins et al. [56])

No	Event	Level 1	Level 2
1	V_{NE} (Never Exceed Velocity)	158 knots	163 knots
2	Vertical g Load	3.0	3.8
3	Vertical g Load - Min	-1.0	-1.52
4	Oil Temperature - Max	–	245 F
5	Oil Pressure - Min	–	20 psi
6	Oil Pressure - Max	–	115 psi
7	RPM - Max	2700 RPM or more for ≥ 1s	2700 RPM or more for > 5s
8	Cylinder Head Temperature - Max		500 F
9	Fuel Quantity - Min	8 gal.	5 gal.
10	Bank Angle	60°	≥ 65°
11	Pitch Attitude (positive)	30°	≥ 35°
12	Pitch Attitude (negative)	-30°	≤ -35°
13	Vertical Speed Magnitude Below 1000 AGL	≥ 800 fpm	≥ 1000 fpm
14	Airspeed at or below 200 feet AGL High Speed Full Flaps	66 knots for 2s	71 knots for 2s
15	Airspeed at or below 200 feet AGL High Speed Zero Flaps	75 knots for 2s	80 knots for 2s
16	Airspeed at or below 200 feet AGL Low Speed Full Flaps	60 knots for 2s.	≤56 knots for 1s
17	Airspeed at or below 200 feet AGL Low Speed Zero Flaps	69 knots for 2s	≤65 knots for 2s
18	Extended Centerline deviation at 200 feet AGL	2°	3°
19	Glide angle High (Too steep) at 200 feet AGL	4°	5°
20	Glide angle Low (Too shallow) at 200 feet AGL	2°	1°
21	Bank Angle at or below 200 feet AGL	20°	25°
22	Pitch Attitude at Touchdown (High)	10.5°	12°
23	Pitch Attitude at Touchdown (Low)	3°	1°
24	Airspeed at Touchdown (High - Full Flap)	55 knots	60 knots
25	Airspeed at Touchdown (High - No Flap)	63 knots	68 knots

During comparison of exceedance detection with instantaneous anomalies, it is noted that parameters used in all the exceedances from Table 3 might not be available to the anomaly detection algorithm. However, it is expected that violating those exceedance thresholds should still be observable and detectable by the instantaneous anomaly detection. Similarly, data to measure engine management-related exceedances (4, 5, 6, 8, 9) may not be available for all GA flight recorders and these exceedances may not necessarily overlap with instantaneous anomalies completely (unless very serious engine issues). However, since the aim of the method is not necessarily replacing the exceedances but adding a new dimension to the safety analysis, it is valuable to understand where the techniques do not overlap.

The exceedances listed here are used in this work for both the Cessna 172 and Piper Archer as the authors did not have access to any other published references for Piper Archer exceedances. This is one of the limitations of using exceedances versus anomaly detection as exceedances can vary from one aircraft to another or vary for the same aircraft among different operators.

B. Summary of Utilized Energy Metrics

Table 4 Summary of implemented energy metrics and data required for computation

		Can be estimated using		
Metric	**Formula**	**Flight Data**	**Flight Data + Ref. Profile**	**Flight Data + Perf. Model**
Specific Total Energy (STE)	$h + V^2/2g$	✓	✓	✓
Specific Potential Energy (SPE)	h	✓	✓	✓
Specific Kinetic Energy (SKE)	$V^2/2g$	✓	✓	✓
Specific Total Energy Rate (STER)	$\dot{h} + V\dot{V}/g = (T - D)V/W$	✓	✓	✓
Specific Potential Energy Rate (SPER)	$\dot{h} = V \sin\gamma$	✓	✓	✓
Specific Kinetic Energy Rate (SKER)	$V\dot{V}/g$	✓	✓	✓
Potential Flight Path Angle (PFPA)	$\gamma + \dot{V}/g$	✓	✓	✓
Energy Rate Distribution (ERD)	$sign(\frac{SKER}{SPER}) \times \exp(-\lvert\frac{SKER}{SPER}\rvert)$	✓	✓	✓
Specific Total Energy Error (STEE)	$h_{act} - h_{ref} + (V_{act}^2 - V_{ref}^2)/2g$	✗	✓	✗
Specific Potential Energy Error (SPEE)	$h_{act} - h_{ref}$	✗	✓	✗
Specific Kinetic Energy Error (SKEE)	$(V_{act}^2 - V_{ref}^2)/2g$	✗	✓	✗
Normalized Specific Energy Error (NSEE)	$((STE)_{act} - (STE)_{ref})/(STE)_{tol}$	✗	✓	✗
Specific Total Energy Error Rate (STEER)	$sign(STEE) \times \delta(STEE)/\delta t$	✗	✓	✗
Inverse Energy Rate Efficiency (IERE)	$V_{act}(T - D)/V_{red}W(\gamma_{ref} + \dot{V_{ref}}/g)$	✗	✓	✗
Max. Potential Flight Path Angle ($PFPA_{max}$)	$T_{max} - D/W$	✗	✗	✓
Min. Potential Flight Path Angle ($PFPA_{min}$)	$T_{idle} - D/W$	✗	✗	✓
Thrust Margin (TM)	$1 - T/T_{max}$	✗	✗	✓
Energy Rate Margin (ERM)	$(W(\gamma_a + \dot{V_a}/g))/(T_{max} - D)$	✗	✗	✓
Energy Rate Demand (ERD)	$W(\gamma_c + \dot{V_c}/g)/(T_{idle} - D)$	✗	✗	✓

The energy metrics defined in the table typically involve using the true airspeed rather than ground speed so as to ensure that the entire kinetic energy rate gets captured. Additionally, all GA aircraft may not have reliable GPS recording capability required to compute ground speed.

C. Minimum Set of Parameters Required for Method

The following table summarizes the minimum set of parameters required for the instantaneous anomaly detection method outlined in this paper.

Table 5 Summary of parameters required for instantaneous anomaly detection method and typical data recording capability of high-end GA flight data recorders

No	Parameter	Typical high-end GA FDR	Energy-based Metric Features
1	Altitude	✓	✓
2	True Airspeed	✓	✓
3	Indicated Speed	✓	✓
4	Vertical Airspeed	✓	✓
5	Pitch	✓	✓
6	Flight Path Angle	✓	✓
7	Angle of Attack	✓	✓
8	Outside Air Temperature	✓	✓
9	Exhaust Gas Temp.	✓	✓
10	RPM	✓	✓
11	Fuel Flow Rate	✓	✓
12	Latitude	✓	
13	Longitude	✓	
14	Altitude (GPS)	✓	
15	Ground speed	✓	
16	Roll	✓	
17	Lateral Acceleration	✓	
18	Normal Acceleration	✓	
19	Heading	✓	
20	Track	✓	
21	Fuel Quantity (Left)	✓	
22	Fuel Quantity (Right)	✓	
23	Oil Temperature	✓	
24	Oil Pressure	✓	
25	Cylinder Head Temp.	✓	

References

[1] "National Transportation Safety Board Accident Statistics for 2013," http://www.ntsb.gov/investigations/data/Pages/AviationDataStats.aspx, 2019. Retrieved: 10/2019.

[2] "Statistical Summary of Commercial Jet Airplane Accidents - Boeing Commercial Airplanes," http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf, 2017. Retrieved: 10/2019.

[3] "National Transportation Research Board Most Wanted List 2017-18," https://www.ntsb.gov/safety/mwl/Documents/2017-18/MWL-Brochure2017-18.pdf, 2018. Retrieved: 10/2019.

[4] "Federal Aviaition Administration Aerospace Forecasts Fiscal Years 2016-2036." https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2016-36_FAA_Aerospace_Forecast.pdf, 2017. Retrieved: 10/2019.

[5] Logan, T. J., "Error Prevention as Developed in Airlines," *International Journal of Radiation Oncology*Biology*Physics*, Vol. 71, No. 1, 2008, pp. S178–S181. doi:10.1016/j.ijrobp.2007.09.040.

[6] "Federal Aviation Administration Advisory Circular, 120-82 - Flight Operational Quality Assurance," https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/23227, 2004. Retrieved: 10/2019.

[7] "Federal Aviation Administration - Aviation Safety Information Analysis and Sharing (ASIAS)," http://www.asias.faa.gov, 2017. Retrieved: 10/2019.

[8] "General Aviaition Joint Steering Committee (GAJSC)," http://www.gajsc.org/, 2017. Retrieved: 10/2019.

[9] "Civil Aviation Authority - Flight Data Monitoring CAP 739 Second Edition," https://publicapps.caa.co.uk/docs/33/CAP739.pdf, 2013. ISBN 978-0-11792-840-4. Retrieved: 10/2019.

[10] Kuo, B. C., Guan, W.-l., and Chen, P.-c., "In Search of General Aviation Flight Data Monitoring: Lightweight Recording System," *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. doi:10.2514/6.2017-3439.

[11] "24th Joseph T. Nall Report - Aircraft Owners and Pilots Association," https://www.aopa.org/-/media/files/aopa/home/pilot-resources/safety-and-proficiency/accident-analysis/nall-report/15-fn-0022-1-24th-nall-v6.pdf, 2012. Retrieved: 10/2019.

[12] "Federal Aviation Administration, 14 CFR §23.2005 Certification of Normal Category Airplanes," https://www.ecfr.gov/cgi-bin/text-idx?node=pt14.1.23#se14.1.23_12005, 2017. Retrieved: 10/2019.

[13] Das, S., Matthews, B. L., Srivastava, A. N., and Oza, N. C., "Multiple Kernel Learning for Heterogeneous Anomaly Detection: Algorithm and Aviation Safety Case Study," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2010, pp. 47–56. doi:10.1145/1835804.1835813.

[14] Matthews, B., Das, S., Bhaduri, K., Das, K., Martin, R., and Oza, N., "Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms," *Journal of Aerospace Information Systems*, Vol. 10, No. 10, 2013, pp. 467–475. doi:10.2514/1.I010080.

[15] Li, L., Das, S., John Hansman, R., Palacios, R., and Srivastava, A. N., "Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations," *Journal of Aerospace Information Systems*, Vol. 12, No. 9, 2015, pp. 587–598. doi:10.2514/1.I010329.

[16] Chandola, V., Banerjee, A., and Kumar, V., "Anomaly Detection: A Survey," *ACM computing surveys (CSUR)*, Vol. 41, No. 3, 2009. doi:10.1145/1541880.1541882.

[17] Gavrilovski, A., Jimenez, H., Mavris, D. N., Rao, A. H., Shin, S., Hwang, I., and Marais, K., "Challenges and Opportunities in Flight Data Mining: A Review of the State of the Art," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, 2016. doi:10.2514/6.2016-0923, paper No: AIAA 2016-0923.

[18] Iverson, D. L., "Inductive System Health Monitoring," Tech. rep., National Aeronautics and Space Administration, 2004. Url: https://ntrs.nasa.gov/search.jsp?R=20040068062.

[19] Bay, S. D., and Schwabacher, M., "Mining Distance-based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2003, pp. 29–38. doi:10.1145/956750.956758.

[20] Amidan, B. G., and Ferryman, T. A., "APMS SVD Methodology and Implementation," Tech. rep., U.S. Department of Energy PNWD-3026, 2000. doi:10.2172/753847.

[21] Mugtussids, I. B., "Flight Data Processing Techniques to Identify Unusual Events," Ph.D. thesis, Virginia Polytechnic Institute and State University, 2000. Url: `http://hdl.handle.net/10919/28095`.

[22] Melnyk, I., Matthews, B., Valizadegan, H., Banerjee, A., and Oza, N., "Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems," *Journal of Aerospace Information Systems*, 2016, pp. 161–173. doi:10.2514/1.I010394.

[23] Li, L., Hansman, R. J., Palacios, R., and Welsch, R., "Anomaly Detection via a Gaussian Mixture Model for Flight Operation and Safety Monitoring," *Transportation Research Part C: Emerging Technologies*, Vol. 64, 2016, pp. 45–57. doi:10.1016/j.trc.2016.01.007.

[24] Keogh, E., Lin, J., and Fu, A., "HOT SAX: Efficiently finding the most unusual time series subsequence," *Fifth IEEE International Conference on Data Mining (ICDM'05)*, IEEE, 2005, p. 8 pp. doi:10.1109/ICDM.2005.79.

[25] Deshmukh, R., and Hwang, I., "Incremental-Learning-Based Unsupervised Anomaly Detection Algorithm for Terminal Airspace Operations," *Journal of Aerospace Information Systems*, Vol. 16, No. 9, 2019, pp. 362–384. doi:10.2514/1.I010711.

[26] Khalastchi, E., Kalech, M., Kaminka, G. A., and Lin, R., "Online data-driven anomaly detection in autonomous robots," *Knowledge and Information Systems*, Vol. 43, No. 3, 2015, pp. 657–688. doi:10.1007/s10115-014-0754-y.

[27] Vespe, M., Visentini, I., Bryan, K., and Braca, P., "Unsupervised learning of maritime traffic patterns for anomaly detection," 2012. doi:10.1049/cp.2012.0414.

[28] Chakravarty, P., Zhang, A. M., Jarvis, R., and Kleeman, L., "Anomaly detection and tracking for a patrolling robot," *Australasian Conference on Robotics and Automation (ACRA)*, Citeseer, 2007. URL `http://www.araa.asn.au/acra/acra2007/papers/paper116final.pdf`.

[29] Alheeti, K. M. A., Gruebler, A., and McDonald-Maier, K. D., "An intrusion detection system against malicious attacks on the communication network of driverless cars," *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, IEEE, 2015, pp. 916–921. doi:10.1109/CCNC.2015.7158098.

[30] Rajbahadur, G. K., Malton, A. J., Walenstein, A., and Hassan, A. E., "A survey of anomaly detection for connected vehicle cybersecurity and safety," *2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2018, pp. 421–426. doi:10.1109/IVS.2018.8500383.

[31] van Wyk, F., Wang, Y., Khojandi, A., and Masoud, N., "Real-Time Sensor Anomaly Detection and Identification in Automated Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2019. doi:10.1109/TITS.2019.2906038.

[32] Chen, Q., Qiu, Q., Wu, Q., Bishop, M., and Barnell, M., "A confabulation model for abnormal vehicle events detection in wide-area traffic monitoring," *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, IEEE, 2014, pp. 216–222. doi:10.1109/CogSIMA.2014.6816565.

[33] Puranik, T., Jimenez, H., and Mavris, D., "Energy-Based Metrics for Safety Analysis of General Aviation Operations," *Journal of Aircraft*, Vol. 54, No. 6, 2017, pp. 2285–2297. doi:10.2514/1.C034196.

[34] Puranik, T., Harrison, E., Min, S., Jimenez, H., and Mavris, D., "Energy-Based Metrics for General Aviation Flight Data Record Analysis," *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016. Paper No. AIAA 2016-3915, doi:10.2514/6.2016-3915.

[35] Harrison, E., Min, S., Jimenez, H., and Mavris, D. N., "Implementataion and Validation of an Internal Combustion Engine and Propeller Model for General Aviation Aircraft Performance Studies," *15th AIAA Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2015. Paper No. AIAA-2015-2850, doi:10.2514/6.2015-2850.

[36] Min, S., Harrison, E., Jimenez, H., and Mavris, D. N., "Development of Aerodynamic Modeling and Calibration Methods for General Aviation Aircraft Performance Analysis - a Survey and Comparison of Models," *AIAA Aviation*, American Institute of Aeronautics and Astronautics, 2015. Paper No. AIAA 2015-2853, doi:10.2514/6.2015-2853.

[37] Puranik, T., Harrison, E., Min, S., Jimenez, H., and Mavris, D., "General Aviation Approach and Landing Analysis Using Flight Data Records," *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016. Paper No. AIAA 2016-3913, doi:10.2514/6.2016-3913.

[38] Puranik, T. G., Jimenez, H., and Mavris, D. N., "Utilizing Energy Metrics and Clustering Techniques to Identify Anomalous General Aviation Operations," *AIAA SciTech Forum*, 2017. Paper No. AIAA 2017-0789, doi:10.2514/6.2017-0789.

[39] Puranik, T., and Mavris, D., "Anomaly Detection in General-Aviation Operations Using Energy Metrics and Flight-Data Records," *Journal of Aerospace Information Systems*, Vol. 15, No. 1, 2018, pp. 22–35. doi:10.2514/1.I010582.

[40] Puranik, T. G., "A Methodology for Quantitative Data-driven Safety Assessment for General Aviation," Ph.D. thesis, Georgia Institute of Technology, 2018. URL `http://hdl.handle.net/1853/59905`.

[41] Rao, A. H., and Puranik, T. G., "Retrospective Analysis of Approach Stability in General Aviation Operations," *18th AIAA Aviation, Technology, Integration, and Operations Conference, Atlanta, GA. June*, 2018. doi:10.2514/6.2018-3049.

[42] Puranik, T. G., Harrison, E. H., Min, S. M., Chakraborty, I., and Mavris, D. N., "A Framework for General Aviation Aircraft Performance Model Calibration and Validation," *18th AIAA Aviation, Technology, Integration, and Operations Conference, Atlanta, GA. June*, 2018. doi:10.2514/6.2018-3191.

[43] Hochreiter, S., and Schmidhuber, J., "Long short-term memory," *Neural computation*, Vol. 9, No. 8, 1997, pp. 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[44] Ochoa, C. A., and Atkins, E. M., "Multicopter Failure Diagnosis through Supervised Learning and Statistical Trajectory Prediction," *AIAA Information Systems-AIAA Infotech@Aerospace*, 2018. Paper Number: AIAA 2018-1636, doi:10.2514/6.2018-1636.

[45] "Federal Aviation Administration, 14 CFR §121.344 Digital Flight Data Recorders for Transport Category Airplanes," `https://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title14/14cfr121_main_02.tpl`, 2011.

[46] Campbell, N., "Flight Data Analysis - An Airline Perspective," *Australian and New Zealand Societies of Air Safety Investigators Conference*, 2003. URL `https://www.asasi.org/papers/2003/Flight%20Data%20Analysis_Campbell_V2.pdf`.

[47] Chakraborty, A., Fala, N., and Marais, K., "Investigation of Feasibility of Using Low-Cost AHRS Devices to Detect General Aviation Hazardous States," *AIAA Aviation 2019 Forum*, 2019. doi:10.2514/6.2019-3444, Paper Number: AIAA-2019-3444.

[48] Nothem, M., "System Identification of a General Aviation Aircraft Using a Personal Electronic Device," Master's thesis, Georgia Institute of Technology, 2019. URL `https://smartech.gatech.edu/handle/1853/61306`.

[49] "Stratus by Appareo," `https://www.foreflight.com/products/stratus/`, 2017. Retrieved: 10/2019.

[50] Garmin, *G1000 Integrated Flight Deck Pilot's Guide*, 2017. URL `http://static.garmincdn.com/pumac/190-00498-07_0A_Web.pdf`, Retrieved 10/2019.

[51] Chandola, V., "Anomaly Detection for Symbolic Sequences and Time Series Data," Ph.D. thesis, University of Minnesota, 2009. Url: `http://hdl.handle.net/11299/56597`.

[52] Anon., *Information Manual: Cessna Skyhawk SP, 1st ed., Pilot Operating Handbook, Model 172S NAV III GFC 700 AFCS, Cessna Aircraft Company*, 2007.

[53] Min, S., "A Proactive Safety Enhancement Methodology for General Aviation Using a Synthesis of Aircraft Performance Models and Flight Data Analysis," Ph.D. thesis, Georgia Institute of Technology, 2018. URL `http://hdl.handle.net/1853/60807`.

[54] Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, 1977, pp. 1–38. URL `http://www.jstor.org/stable/2984875`.

[55] "Flight Safety Foundation ALAR Briefing Note 4.2 - Energy Management," Technical Report, 2000. URL `https://flightsafety.org/files/alar_bn4-2-energymgmt.pdf`, Retrieved: 10/2019.

[56] Higgins, J., Clachar, S., and Hennselek, K., "Flight Data Monitoring (FDM)/Flight Operational Quality Assurance (FOQA) for General Aviation," Tech. rep., U.S. Department of Transportation - Federal Aviation Administration, December 2013. Technical Report DOT/FAA/AR-13/21.